Scikit Learn and PyTorch Installation Guide

Marios S. Pattichis
Dept of Electrical and Computer Engineering
The University of New Mexico
pattichi@unm.edu
ECE 229A
(505) 277-0486

Fall 2020

1 Install Anaconda

To install the free version of `Anaconda`, it is recommended for to download the latest version identified as: Anaconda individual licence based on `Python 3.8` and a `64-Bit Graphical Installer`.

2 Required: Freely availabe books and required book example downloads

We have Python two required books:
1. *Hands-On Machine Learning with Schikit-Learn & Tensorflow*, 2nd edition.
2. *Programming PyTorch for Deep Learning: Creating and Deploying Deep Learning Applications*

Both books are freely available to all UNM students. To access them, simply go to `https://libguides.unm.edu/Safari`. There are several other Python textbooks available through this link.

The books come with github accounts that provide the code examples. You will need to download the examples from:
1. For the Hands-on book refer to Hands-on github .
2. For the PyTorch book refer to PyTorch github.

3 Installation of PyTorch after installing Anaconda

The following installation uses `conda` commands. You can download the Anaconda environment cheatsheet from `https://conda.io/docs/_downloads/conda-cheatsheet.pdf`.

Run `Anaconda Prompt (Anaconda 3)` as Administrator. Then, in the prompt, type:

```
(base) ...>conda create --name pytorch
(base) ...>conda activate pytorch
(pytorch) ...>conda install pytorch torchvision cudatoolkit=10.2 -c pytorch
(pytorch) ...>conda install -c conda-forge jupyterlab
(pytorch) ...>conda install pandas
(pytorch) ...>conda install -c conda-forge matplotlib
```

You need to install the DLLs for Microsoft Visual C++ and restart your computer. This is accomplished by:

```
1. Open a browser and type https://aka.ms/vs/16/release/vc_redist.x64.exe
2. Download vc_redist.x64.exe in the Download directory and
3. Run vc_redist.x64.exe as an administrator
4. Retart the Computer
```

Test the installation using:

```
(pytorch) ...>python
>>>import torch
>>>
>>>exit()
(pytorch) ...>cd to-your-Python-notebook-directory
(pytorch) ...>jupyter notebook
```

3.1 Spyder: An Interactive Python Editor

Spyder allows us to run interactive examples. As before, run Anaconda Prompt as an Administrator and do:

```
(base) ...>conda activate pytorch
(pytorch) ...>conda install spyder
spyder
```

3.2 Optional: Running the Natural Language Processing text examples

If you want to add the modules for Natural Language Processing support, then run `Anaconda Prompt (Anaconda 3)` as Administrator, and follow the steps to download the English Natural Language models:

```
(base) ...>conda activate pytorch
(pytorch) ...>conda install -c pytorch torchtext
(pytorch) ...>conda install -c conda-forge spacy
(pytorch) ...>python -m spacy download en_core_web_sm
(pytorch) ...>conda install -c conda-forge googletrans
```

Note: To add Spanish, you will need to modify the first function call. After you complete these steps, you should be able to run the examples of Chapter 5 of the book.

3.3 Optional: Running the audio processing examples

Similarly, if you want to install support for processing sound examples, then download the `ESC-50` dataset from ESC-50 download link. You can then install

```
(pytorch) ...>conda install -c conda-forge librosa
(pytorch) ...>conda install -c pytorch torchaudio
(pytorch) ...>pip install PySoundFile
```

After you complete these steps, you should be able to run the examples of Chapter 6 of the book. **Please note that the spectrogram examples require a lot of time to be pre-computed.**

3.4 Running PyTorch Book Examples

To run the book examples, you will need to first download the Jupyter notebooks from github for Programming PyTorch for Deep Learning. Then, after downloading the examples, you will need to navigate to the book examples directory and run the following:

```
(base) ...>conda activate pytorch
(pytorch) ...>python setup.py
(pytorch) ...>mkdir tmp
(pytorch) ...>jupyter notebook
```

Ignore any errors associated with not being able to find certain files. The system will work with the images that it does find.

Then open up `Chapter 2.ipynb` and look for the `tmp` directory and remove the leading \. For example, modify the directory \tmp\simplenet to tmp\simplenet.

3.5 Running PyTorch after installation

Every time you need to login, simply follow:

```
Run the Anconda Prompt.
Then switch to the pytorch environment using:
(base) ...>conda activate pytorch
(pytorch) ...>jupyter notebook
```

3.6 Remove Anaconda Environment if something goes wrong

Make aure to restart Anaconda Prompt as Administrator. Then type:

```
conda env remove --name pytorch
```

You can then restart the installation process. You will get a warning that the pytorch directory exists. This is fine. You can overwrite it.

4 References

4.1 Jupyter notebook basics

Saving the notebook does not save the values of Anaconda variables. It only saves the printed output text. You will need to rerun the code to reproduce the same environment with your saved values.

If you are interested in using markdown to document your Jupyter notebooks, you can study the basics at: `https://daringfireball.net/projects/markdown/basics`.

4.1.1 Start Jupyter notebook

Open up a new Anaconda Prompt and type:

```
conda activate pytorch
jupyter notebook
```

4.1.2 Basic operations

- **Execute a Cell:** Click `Run` or simply hit `Shift-Enter` on the Cell.
- **Run all from scratch:** Click `Cell` and then `Run All`.
- **Save the current notebook:** Click on the disk icon to save your work.
- **Interrupt long computation:** Click on `Kernel` and then click on `Interrupt`. This approach KEEPS ALL VARIABLES. Restart running after Kernel interrupt
- **To run the rest do:** Click on Cell and select `Run All Below`.

4.1.3 Restart up to a point

```
To run up to a point: Click inside the last Cell
Click on Kernel and Restart to lose all variables.
Click on Cell and Run All Above.
```

4.1.4 Exit after all computations are complete

This process will loose all of your variables. You can only preserve the current outputs.

```
Click on the disk icon to save your work.
Click Kernel then Shutdown.
Close the browser windows.
Then close Anaconda Prompt.
```

4.1.5 Emergency exit if Interrupt does not work after long wait

This process will loose all of your variables. You can only preserve the current outputs:

```
Save the current notebook by clicking on the disk icon.
Simply type Control-C in the Anaconda Prompt.
You may have to press Control-C multiple times.
You can then click and close each tab in the browser.
```

4.2 Scikit Learn Software Examples

Scikit learn has excellent packages for machine learning that we will be using. Refer to: `https://scikit-learn.org/stable/`.

Clustering A very nice example that shows many clustering algorithms can be found at `https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html`.

Classifiers A very nice example that compares several classifiers can be found at `https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html`.

Face classification in the wild example The example uses PCA and SVM to recognize faces as given in `https://scikit-learn.org/stable/auto_examples/applications/plot_face_recognition.html`.

Prediction latency This is important for large datasets. See `https://scikit-learn.org/stable/auto_examples/applications/plot_prediction_latency.html`.

4.3 PyTorch Tutorials

Refer to the PyTorch Tutorials Website for the latest information.

- PyTorch Starting Examples
- Pruning the size of a Neural Network
- PyTorch Cheatsheet

4.4 PyTorch Resources

Please refer to `Safari` under UNM Library resources for Computer Science.

You can create an account with your UNM ID in O Reilly website to gain full access to `Programming PyTorch for Deep Learning`.

4.5 Loading Model Languages in Spacy

Refer to Spacy Language Models to load different models.